ehp ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

A Unified Probabilistic Framework for Dose–Response Assessment of Human Health Effects

Weihsueh A. Chiu and Wout Slob

http://dx.doi.org/10.1289/ehp.1409385

Received: 24 October 2014
Accepted: 19 May 2015
Advance Publication: 22 May 2015

This article will be available in a 508-conformant form upon final publication. If you require a 508-conformant version before then, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

NIH National Institute of Environmental Health Sciences

Environ Health Perspect DOI: 10.1289/ehp.1409385
Advance Publication: Not Copyedited

A Unified Probabilistic Framework for Dose–Response Assessment of Human Health Effects

Weihsueh A. Chiu[1*] and Wout Slob[2]

[1]National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC, USA; [2]National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands

***Present address:** Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA

Address correspondence to Weihsueh A. Chiu, Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, VMA Building, Rm. 107, 4458 TAMU, College Station, TX 77843-4458 USA. Telephone: 979-845-4106. E-mail: wchiu@cvm.tamu.edu

Running title: Unified probabilistic dose–response assessment

Acknowledgments: The authors would like to acknowledge the other Working Group members of the World Health Organization/International Program on Chemical Safety Project on Uncertainty in Hazard Characterization for their contributions to the development of this approach, including Bernard Bottex, David Bussard, George Fotakis, Andy Hart, Dale Hattis, Matthias Herzler, Kathy Hughes, Theo Vermeire, Carolyn Vickers, Paul White, and Lauren Zeise. In addition, R. Woodrow Setzer, Kan Shao, Ted Berner, Steven Bradbury, and Tony Cox, Jr. provided helpful comments that substantially improved this manuscript.

Disclaimer: The views expressed herein do not necessarily represent those of the U.S. Environmental Protection Agency.

Competing financial interests: None of the authors has any actual or potential competing financial interests.

Abstract

Background: When chemical health hazards have been identified, probabilistic dose-response assessment ("hazard characterization") quantifies uncertainty and/or variability in toxicity as a function of human exposure. Existing probabilistic approaches differ for different types of endpoints or modes-of-action, lacking a unifying framework.

Objectives: Developing a unified framework for probabilistic dose-response assessment.

Methods: We establish a framework based on four principles: (1) individual and population dose-response are distinct; (2) dose-response relationships for all (including quantal) endpoints can be recast as relating to an underlying continuous measure of response at the individual level; (3) for effects relevant to humans, "effect metrics" can be specified to define "toxicologically equivalent" sizes for this underlying individual response; and (4) dose-response assessment requires making adjustments and accounting for uncertainty and variability. We then derive a step-by-step probabilistic approach for dose-response assessment of animal toxicology data similar to how non-probabilistic reference doses are derived, illustrating the approach with example noncancer and cancer datasets.

Results: Probabilistically-derived exposure limits are based on estimating a "target human dose" HD_M^I, which requires risk management-informed choices for the magnitude (M) of individual effect being protected against, the remaining incidence (I) of individuals with effects $\geq M$ in the population, and the percent confidence. In the example datasets, probabilistically-derived 90% confidence intervals span a 40-60-fold range for HD_M^Is where I=1% of the population experiences $\geq M$=1%-10% effect sizes.

Conclusions: While some implementation challenges remain, this unified probabilistic framework can provide substantially more complete and transparent characterization of chemical hazards and support better-informed risk management decisions.

Introduction

While the process of identifying human health hazards of chemicals has evolved substantially over time with advances in weight of evidence determination, mode of action (MOA), and systematic review (e.g., Meek et al. 2013; NRC 2011; U.S. EPA 2005; Woodruff and Sutton 2014), practices for quantitative dose-response assessment to characterize those hazards and inform risk management rely largely on approaches that have shown relatively small changes since they were first used. For assessment of non-cancer effects, it is still common to derive exposure limits by dividing a no observed adverse effect level (NOAEL) or a benchmark dose lower confidence limit (BMDL) derived from a chronic study by a generic "uncertainty factor" of 100 (also known as a "safety," "assessment," or "extrapolation" factor), although using chemical-specific "adjustment" factors (CSAFs) or data-derived extrapolation factors (DDEFs) is increasingly encouraged (IPCS 2005; U.S. EPA 2014a). Exposure limits for carcinogens that are genotoxic or without an established non-genotoxic MOA are usually based on other approaches, in particular the linear extrapolation approach (EFSA 2005; U.S. EPA 2005), although more recently the margin-of-exposure approach has been suggested even for genotoxic carcinogens (Barlow et al. 2006; Benford et al. 2010; O'Brien et al. 2006) .

While procedurally straight-forward, these practices are most amenable to risk management decisions where the margins between calculated exposure limits and actual or anticipated exposures are large enough to be of little or no risk management concern. For instance, the safety factor approach results in an exposure limit (acceptable daily intake [ADI], reference dose [RfD]) generally presumed to be "safe" (e.g., having "reasonable certainty of no harm"). However, the conclusion that exposures at or below this level would not result in

appreciable health risks is typically not based on further quantitative substantiation. For exposures higher than such an exposure limit, the only statement that can be made is that risks "cannot be excluded," without any quantitative characterization of what the extent of potential health effects might be. Thus, if reducing exposures to the level of the derived exposure limit is challenging (e.g., economically, practically, or politically), then there is no way to weigh the cost of exposure reduction against its likely human health benefits. Moreover, there may be a residual risk at, or even below, the exposure limit, and this residual risk may vary among different chemicals and/or exposure scenarios.

To address these disadvantages, a probabilistic approach to hazard or risk characterization has been advocated by several risk assessment researchers (Baird et al. 1996; Evans et al. 2001; Hattis et al. 2002; Slob and Pieters 1998; Swartout et al. 1998), as well as by several expert panels (NRC 1996; NRC 2009; SAB 2002). While most of the work has focused on characterization of non-cancer effects, NRC (2009) revisited the question of unifying assessment of cancer and non-cancer effects. While some of the recommendations of NRC (2009) as to default approaches to low-dose extrapolation have been controversial (Abt et al. 2010; Pottenger et al. 2011; Rhomberg 2011; Rhomberg et al. 2011), deciding on such science policy questions (as "default" options clearly are) does not preclude moving forward with developing a unified probabilistic framework for all types of endpoints.

This paper presents such a unified probabilistic framework, developed in tandem with an international harmonization project (IPCS 2014) on uncertainty in human dose-response assessment ("hazard characterization" in WHO/IPCS nomenclature). It retains the usual "two-part" process as employed in the current non-probabilistic ("deterministic") approaches for

quantitative dose-response assessment: (1) dose-response analysis of an experimental or observational dataset of health effects resulting from chemical exposure; and (2) inference (or "extrapolation") as to the potential effects in the target human population. The second part needs to account for differences in characteristics (e.g., species, exposure duration) between the dataset analyzed and the human population of interest for risk assessment. In terms of the usual deterministic approach of dose-response assessment, the determination of the point of departure (POD) may be regarded as the first part and the extrapolations addressed by uncertainty factors (interspecies, intraspecies, subchronic-to-chronic, etc.) as the second part. Although the basic procedure appears unchanged, probabilistic assessment requires more precise definitions of each step, and thus also promotes greater transparency as to the biological and quantitative assumptions underlying the dose-response assessment. In particular, the framework we propose here provides a theoretical basis for human dose-response assessment, where all underlying concepts are explicitly defined and logically interrelated. In this way, it is fully transparent as to what the various computational procedures represent, and how the results can be interpreted.

This paper is organized as follows. In "Methods," we first set out the four fundamental principles that underlie the unified probabilistic framework. Additionally, we lay out a prototypical approach to implement the unified framework for human-relevant animal toxicology data. In the "Results," we illustrate the approach by deriving probabilistic exposure limits from example noncancer and cancer datasets using probability distributions for uncertainty derived from historical data (see Supplemental Materials, Table S1, for datasets and computer code). In "Conclusions," we discuss implementation issues and identify data needs (also see Supplemental Material, Additional applications and extensions).

Methods

Fundamental Principles

Principle 1. Individual and population-level dose-response

The starting point of this framework is that a conceptual distinction exists between effects on the individual and effects on the population. In particular, the effect of exposure at the level of the individual is the ***magnitude*** of a measure of toxicological effect. The result of a fixed exposure in a population will be different magnitudes of effect in different individuals in that population. Therefore, for a particular magnitude of effect, the result in the population is expressed as an ***incidence***. In the present framework, the magnitude of change needs to be ordinally related to severity – so a greater magnitude constitutes a more severe effect. For instance, a body weight (BW) decrease of 20% is more severe than a BW decrease of 10%, and a moderate liver lesion is more severe than a mild liver lesion. Thus, for a monotonic dose-response in an individual, it may be imagined that a higher exposure will, for any given endpoint, lead to more severe effects. In a population, increasing exposure levels will result in simultaneously increasing both incidence and severity: more and more individuals will suffer from more and more severe effects.

For convenience, we establish the notation whereby human dose or exposure is denoted *HD*, the magnitude of effect is denoted by *M*, and incidence is denoted *I*. Because *M* is assumed to have an ordinal relationship with severity, incidence can be characterized as the incidence of effects of magnitude *equal or greater than M*, denoted $I_{\geq M}$. Because it is customary to discuss "incidence" in terms of effects that may be of concern, the simpler notation HD_M^I will be used as shorthand for $HD(I_{\geq M})$. Additionally, an asterisk (*) will be used to indicate fixed, or target

values, such as a "critical effect size" (M^*), target human dose (HD^*), or target incidence level (I^*).

Given these definitions, the output of a human dose-response assessment is concerned with the quantitative relationships among HD, M, and I, along with their uncertainty. We focus on the most common type of output, which is developing a human health-based exposure limit (some other types of outputs are discussed in Supplemental Materials, Performing a population assessment). In our notation, this means estimating a target human dose, HD^*, which is regarded as a function of a two-dimensional protection goal: the target level of incidence (I^*) and the specified level of effect magnitude (M^*), both of which may be selected based on risk management considerations. Specifically, it is the dose at which only a small fraction of the population (low incidence of I^*) will experience effects $\geq M^*$, which can be written

$$HD^* = HD(I^*_{\geq M^*}) = HD_{M^*}^{I^*}. \quad (1)$$

For instance, one could write $HD(1\%_{\geq 10\%\mathrm{BW}}) = HD_{10}^{01}$ for the dose at which only 1% of the population has greater than 10% change in BW.

Principle 2. Continuous parameters underlying all observed dose-response endpoints

The second element of this framework is that observed dose-response relationships for all endpoints can, at the individual level, be recast as relating to an underlying continuous measure of response. Obviously, this principle applies to endpoints that are directly observed as continuous data. In that case, the observed dose-response for the average responses as a function of dose may be imagined to reflect the dose-response in an individual animal (*viz.* the average animal), even though an individual's dose-response is not directly observable in most

toxicological studies. Endpoints that are generally measured as quantal response rates in a study population require some additional discussion. Below we discuss two options, briefly indicated as "deterministic" and as "stochastic" quantal endpoints (Slob et al 2014).

Many quantal endpoints, such as in histopathology, are ordinally scored (e.g., minimal, mild, moderate, severe). Such endpoints can be considered as gradually increasing in magnitude at the individual level, but are reported in "bins", or severity categories, rather than as a continuous measure. In this way, the reported incidences can be thought of as relating to a single category (or a limited number of categories) of severity, while for other severities the incidences are not reported. In fact, any continuous dataset can be "quantized" and transformed into such a quantal (or ordinal) dataset by setting one (or more) cut points, resulting in incidences Y associated with each cut-point. For instance, changes in hematocrit from the mean level in the controls can be divided into <5% and >5%, and the fraction of animals below and above the 5% cut-point can treated as quantal data. In this case, the effect dose for a 5% change in the continuous hematocrit data, i.e., ED_{05}, is equal to the effect dose for a 50% quantal response, $ED_{Y=50\%}$, as illustrated in Figure 1. This concept of quantal endpoints has previously been discussed in Slob and Pieters (1998).

Therefore, when quantal data can be viewed as reflecting the incidence of a continuous effect above/below a "determined" cut-point equal to M^*, then the endpoint is referred to as a ***deterministic quantal endpoint***. This is generally appropriate for effects that can occur in different degrees of severity. Furthermore, for the purposes of dose-response data analysis, the ED_{50} from the quantal response data would be used to estimate the ED_{M^*} corresponding to M^* of the underlying continuous data. When the available dose-response data report the incidences

related to various severity categories, then one of them may be chosen as being minimally adverse. When they only report the incidences related to a single severity category, it may occur that this severity is more than minimally adverse, so that additional uncertainty arises.

However, not all quantal effects may be derived from applying a cut-point to an underlying continuous variable. Some effects appear to have discrete outcomes, without an underlying gradually increasing level of severity. An example of such an endpoint is malformations, which often do not show different degrees of severity: it is there or it is not. Cancer may be considered another example, since a particular tumor is present or not (ignoring observational practicalities). For such endpoints, an alternative interpretation of the dose-response is possible: the observed incidences at each dose are considered as resulting from a "stochastic" process, where the observation that an individual animal has a tumor or not is analogous to drawing a lottery-ticket, with probability equal to the expected incidence at that dose (and time of observation). That is, given all relevant circumstances for the particular individual (such as genetic make-up, experimental conditions), the effect is not fully determined, but rather any particular animal may be (un)lucky or not. If it were possible to perform a study in which all animals were identical, and identically treated (except the dose, but without dosing errors), then the quantal dose-response would estimate the ***individual probability of effect***. In this case, the observed incidence Y is treated as an estimate of the underlying individual probability of effect M, so M^* would correspond to an incidence $Y^*=M^*$, as depicted in Figure 2. This concept of quantal endpoints has previously been discussed in Slob et al. (2014).

Therefore, when quantal data are assumed to reflect the individual probability of an outcome as a result of a stochastic process, then the endpoint is referred to as a ***stochastic***

quantal endpoint. In reality, there are always small differences between animals (including the experimental conditions), which will have some impact on the dose-response. However, this additional impact is generally not separately distinguishable from the dose-response data. This inter-individual variability might be assumed to be relatively small in the study population and hence ignored. If so, the observed quantal dose-response approximates the individual probability of effect as a function of dose.

Whether the "deterministic" or "stochastic" interpretation of quantal endpoints is correct remains uncertain, since even for endpoints such as cancer or malformations, it might be the case that the effect in an individual subject is evoked deterministically as soon as a given internal dose in that individual is reached. For risk assessment, the distinction between the two interpretations is important for the following reason. In the deterministic interpretation, the animal dose-response curve reflects the experimental variation and errors in the animal study, and therefore its shape (e.g., slope) *would not* be relevant information for predicting risks in humans. In the stochastic interpretation, the dose-response curve may be regarded as a model for the human individual probability of effect, and therefore its shape *would* be relevant as information for human risks.

Principle 3. Selecting a basis for inference: the "effect metric"

The third fundamental element in this framework is that inferences are made on the basis of a selected ***effect metric*** that defines "toxicological equivalent" magnitudes of effect. This effect metric should reflect the effect size in such a way that it applies across species (or populations) as well as across individuals within a species (or population). Changes of the same magnitude in this metric are considered to reflect equal toxicologically-induced changes.

Additionally, the magnitude of effect should also be ordinally related to severity at the level of the individual. For a continuous endpoint, severity increases with an increase in the percent change of a continuous endpoint (e.g., 5% → 7% decrease in hematocrit). For a deterministic quantal endpoint the severity is related with the category of effect (e.g., less severe than "mild" lesions → "mild" or more severe liver lesion). For a stochastic quantal endpoint, severity is related to the probability of experiencing the effect (e.g., 1% → 2% individual probability of cancer).

Equipotent doses are therefore *defined* as doses that elicit the same size of effect metric. Thus, individuals with the same equipotent doses (at all effect sizes) are defined as equally sensitive to the chemical for the endpoint.

Note that it is assumed that the effect has already previously been determined to be an appropriate basis for making inferences about human health effects – e.g., that the effects observed in the test animal are relevant to humans. In this context, "relevance" only needs to be determined in the qualitative sense: Could a similar effect occur or in humans, or not? When the answer is yes, quantitative differences, including large differences that are expected based on MOA considerations, should be addressed explicitly and quantitatively, taking uncertainty into account as well (e.g., using a probabilistic CSAF or DDEF – see Supplemental Material, Chemical-specific/data-derived toxicokinetics or toxicodynamics).

The use of an effect metric does not necessarily imply that a given change is equally adverse in all individuals (or species). For instance, a 5% decrease in hematocrit may be considered as a toxicologically-equivalent effect metric in all individuals, but be adverse in

persons with anemia and non-adverse in healthy persons. Finally, it should be noted that further inferences are possible from the toxicologically-equivalent effect metric to other measures of health effect, such as if an adverse outcome pathway can quantify the linkage between a change in effect metric and the likelihood of an adverse health outcome. For instance, if the effect metric is a percent change in serum cholesterol, given an adverse outcome pathway linking serum cholesterol changes to cardiovascular disease, one might aim to estimate the risk of fatal myocardial infarction in a specific population. Because variability in baseline serum cholesterol levels and other relevant risk factors (e.g., blood pressure, C-reactive protein, etc.) may differ across different populations (e.g., geographic regions, socioeconomic groups, lifestages, etc.), analyses of such "downstream effects" would necessarily be specific to the population(s) being assessed, even if the relationship between exposure and the effect metric is assumed to be the same across populations. Such analyses may also be useful for socioeconomic analyses, since a fixed magnitude of effect may have different cost implications across human subpopulations (e.g., modifying insulin for diabetics versus non-diabetics). This is discussed further in Supplemental Material, Extrapolation to downstream health endpoints and adverse outcome pathways and Figure S1.

Principle 4. Making adjustments while accounting for variability and uncertainty

The final fundamental element in this framework is that dose-response assessment involves making inferences about the human population of interest for risk assessment (the "target population") based on information obtained from a scientific study (the "study population"). In the usual deterministic approach, these inferences are accomplished using the "uncertainty factors" to address (potential) differences due to differing species, human

variability, suboptimal study conditions, etc. However, these factors are often mixtures of multiple elements that need to be clearly specified in a probabilistic framework. Specifically, making inferences between the "study" and "target" populations involves making adjustments from the "study" to the "target" populations, while accounting for variability and uncertainty:

- Adjustments are needed to correct for differences between the "study" and "target" population, in order to make inferences as to the potential health effects in the population of interest, with the relevant exposure conditions. For example, on average across chemicals, the dose in mg/day eliciting the same effect differs between species due to differences in body size. The usual (implicit) adjustment is to divide the dose by body weight, which is also intended to normalize across individual subjects in the (study or target) population. But data increasingly support the idea that the dose in mg/kg BW may need additional adjustment by an allometric scaling factor to achieve equivalent effects (e.g., Bokkers and Slob 2007; Dedrick 1973; Kleiber 1932; Price et al. 2008; U.S. EPA 2011a). Further, it might be known that, for any particular chemical, there are specific differences in toxicokinetic or toxicodynamic properties, which, for instance, make it plausible that one species would be more sensitive than others (e.g., resulting in a CSAF or DDEF). As another example, for some classes of effects, the expected relationship between a benchmark dose and duration of exposure might be reflected by Haber's law (toxicity depends on the product of concentration and exposure time), which may be used to adjust the BMD to other exposure durations. Usually, differences in study population/conditions and the target population/conditions can be better characterized (i.e., its uncertainty reduced) with additional data or analysis, and some can even be

eliminated by conducting new studies that require fewer adjustments (e.g., conducting a chronic study to replace a subchronic study).

- Variability refers to intrinsic heterogeneity about a central tendency, usually between the individuals in the "target" population. For example, different individuals (humans) will exhibit different sensitivity to toxic effects from the same exposure due to various sources of variability, including genetics, life-style, health status, etc. Additional data or analysis can make an estimate of human variability more precise, but the variability itself cannot be eliminated.

- Uncertainty refers to a lack of knowledge that could, in principle, be reduced with additional data or analysis. Uncertainty can relate to the degree of adjustment (e.g., the exact allometric power with which to adjust for BW differences) but also to the degree of variability (e.g., how much more sensitive is the 95% individual relative to the median individual). As another example, because toxicity studies have finite numbers of individuals per dose group, the BMD is uncertain. This uncertainty can, in principle, be reduced by performing a larger or better designed study. Similarly, "missing studies" represent an uncertainty that can be quantified by meta-analyses comparing the overall differences between study types and capture that in a distribution (e.g., Hattis et al. 2002). In some cases, observed variability among chemicals in general can be used to inform the uncertainty in an adjustment factor for a specific chemical. For instance, observed variability among chemicals in the dose ratio between subchronic and chronic studies for the same effect translates into uncertainty in the subchronic-chronic difference for a specific chemical for which no such data are available (e.g., Bokkers and Slob 2005).

As Table 1 shows, all typical uncertainty factors include an uncertainty component, along with an adjustment component, except for the intraspecies factor, where the adjustment component is replaced by a variability component.

Prototypical Approach Implementing a Unified Probabilistic Framework

The principles described above underlying a unified probabilistic framework can be applied to any type of study or endpoint that has dose-response information, but here we address the most common case of using animal toxicology data. It is assumed that the candidate critical endpoint(s) from the animal toxicology study is (are) considered relevant in the sense that similar effects might be expected to occur in humans (uncertainty in the qualitative cross-species concordance is not addressed in this framework). The following additional assumptions are then made:

- The toxicity data are from a study conducted in an (inbred) laboratory animal strain, with the purpose of mimicking what might happen in a typical human being. Intra-study variability reflects experimental errors (e.g., dosing errors, imperfectly controlled experimental conditions) and remaining differences (genetic, or otherwise) among animals. This is treated as statistical uncertainty in estimating a POD, which is supposed to mimic an equipotent dose in the typical human being.

- Furthermore, in the effect range of interest, the continuous dose-response relationships are monotonic and parallel on a log-dose scale across species and across individuals within a species, so that the values (distributions) for any adjustments, variability, or

uncertainties are independent of the selected critical effect size M^*. Slob and Setzer (2014) found evidence consistent with this assumption.

The basic steps of the procedure under these assumptions are as follows (see also Figure 3 and Table 2):

1. Select a toxicologically equivalent ***effect metric*** and an associated ***critical effect size*** (M^*), and conduct a benchmark dose analysis with benchmark response (BMR)=M^* (Crump 1984) to derive the uncertainty distribution for the dose corresponding to M^* in the animal $\rightarrow AD_{M^*}$.

2. Apply probabilistic interspecies and other adjustments to AD_{M^*} to derive the uncertainty distribution for the dose corresponding to M^* in the median human $\rightarrow HD_{M^*}$.

3. Select a human variability distribution (e.g., lognormal), setting the median to HD_{M^*} with an uncertainty distribution as obtained in step 2. The measure of dispersion of this human variability distribution (such as geometric standard deviation, GSD = $\exp[\sigma_H]$) has an uncertainty distribution, reflecting that we are uncertain about the degree variability among individuals. From this (uncertain) human variability distribution, we derive an (uncertain) human variability factor HV_{I^*} for ratio between the quantile corresponding to a selected ***target incidence*** (I^*) value and the median, so that $HD_{M^*}{}^{I^*} = HD_{M^*} \times HV_{I^*}$.

This output is an estimate of the $HD_{M^*}{}^{I^*}$ in the form of an uncertainty distribution, and any given level of confidence may be chosen for deriving an exposure limit (e.g., a "probabilistic RfD"), by taking the associated lower percentile of the uncertainty distribution of $HD_{M^*}{}^{I^*}$. Alternatively, the full uncertainty distribution can be combined with exposure information to inform risk

management decisions. Details as to each step are described below along with Monte Carlo procedures for the overall calculation.

Step 1: Estimating the animal dose corresponding to the critical effect size for the selected toxicologically equivalent effect metric

The purpose of this step is to establish the uncertainty distribution for AD_{M*}, the animal dose associated with a specified effect size M* (= BMR) based on a specified toxicologically-equivalent effect metric.

The key issue in defining the effect metric is how to address baseline differences across species or individuals in order to make changes "comparable." For instance, a decrease of 10 g in a rat body weight does not compare to a 10 g change in human body weight. For most (continuous) parameters, a percent change would be the obvious effect metric, being the only measure that may be defined as representing an equal effect size among different species and individuals (with different background responses). Note that an equal effect size does not imply that it will always be equally adverse in different species/individuals (such as a 5% decrease in hematocrit in anemic versus non-anemic persons). Severity categories in histopathological lesions appear to directly apply as a measure of equivalent effect magnitude. But for endpoints measuring an increase in individual probability of effect, the question of how to correct for the background risk is not easily answered. Various measures are being used, such as additional, extra or relative risk, which all correct for background in a different way. It remains unclear, however, which of these measures reflects an equivalent measure of risk (if any), in particular when background risks among species (populations) differ greatly.

After having chosen the effect metric, one also needs to specify a ***critical effect size*** – the magnitude of effect size M^* of interest, as defined by the problem formulation and risk management context for the assessment. The term *critical* here should be understood in a wide sense, that is, it is a selected value (or even a range of values) that forms a starting point for doing the probabilistic calculations. Often, the problem formulation suggests that the critical effect size reflect the effect size that is considered to be "minimally adverse" biologically. However, current toxicological knowledge does not allow one to unequivocally define minimally adverse effect sizes for all potentially critical endpoints. Further, a given effect size might not be minimally adverse in one species or individual while it is minimally adverse in another (e.g., hematocrit and anemia, discussed above). As a practical limitation, the choice of M may be restricted by the available data. For instance, the reported data may relate to discrete values of M only (e.g., specific severity categories of lesions, as in histopathological data). Moreover, the lower the value of M, the less precise the estimates of the associated doses will be. For biologically-defined M^*s, one might aim to specify study designs that are likely to achieve "adequate" statistical precision for dose estimates related to that value M^*. However, even then, the study design needed to achieve that goal may be impractical (e.g. unrealistic number of animals needed). If so, one may decide to use a statistically-based M^* (i.e., the lowest value of M^* that achieves the desired level of statistical precision) as a surrogate. Such statistically-based M^*s could reflect levels of effect that are larger than minimally adverse levels, and this can be regarded as an additional source of uncertainty or addressed by setting a more stringent protection goal in terms of incidence. Typical examples of effect metrics and critical effect sizes are shown in Table 3, along with the benchmark dose (BMD) approach implied, by treating all endpoints as fundamentally continuous.

The result of the dose-response analysis is an uncertainty distribution for AD_{M*}, the animal dose corresponding to M*. Approaches to establishing the uncertainty distribution include (i) translating the BMD confidence limits obtained by BMD software into a distribution, (ii) parametric bootstrapping (Slob and Pieters 1998; implemented in the R package PROAST [RIVM 2012]), or (iii) Bayesian analysis (Kopylev et al. 2007). It should be noted that fitting a single dose-response model may not fully capture the uncertainties in the dose-response data. Therefore, instead of deriving a BMD distribution from a single model, various models should be fitted to address model uncertainty. These model-specific distributions may be simply pooled in a single distribution (e.g., Slob et al. 2014), or one may apply "model averaging," for which various approaches have been proposed (Bailer et al. 2005; Shao and Gift 2013; Wheeler and Bailer 2007). In addition, if different dose-response datasets are available for the same endpoint, they could be combined in a joint dose-response analysis, with study as a covariate in the analysis, i.e., some of the parameters of the dose-response model are study specific, and others are not (Slob and Setzer 2014).

Step 2: Adjustments due to interspecies differences and study conditions

The purpose of this step is to establish an uncertainty distribution for the "typical" human dose associated with a specified magnitude of effect and endpoint, and with specified exposure conditions. This step combines with the results of Step 1. The "typical" human is defined as the median person of the population. This interspecies step involves addressing three separate aspects:

- *DAF*: A dosimetric adjustment factor for generic physiological differences (e.g. body size differences for oral dose; respiratory tract differences for inhalation exposures) between the test animal and (median) human, along with uncertainty in the expected adjustment;
- *AHU*: Animal-to-human uncertainties due to potential chemical-specific toxicokinetic or toxicodynamic differences between the test animal and humans resulting in differences in sensitivity for a given chemical; and
- *OU*: Other uncertainties due to specific study conditions that differ from the target exposure conditions (e.g., exposure duration, or exposure pattern).

The result of this step is an uncertainty distribution for the human dose at which 50% of the human population has effects greater than (or equal to) M^*:

$$HD(0.5_{\geq M^*}) = AD_{M^*} \times DAF / (AHU \times OU) \quad (2)$$

Each of the adjustments is described in more detail below.

Dosimetric adjustments

It is increasingly evident that generic differences in physiology (e.g., body size) across species can be accounted for by multiplying the animal dose by a dosimetric adjustment factor (*DAF*), or equivalently, by dividing by an "assessment" factor accounting for interspecies body size differences ($AF_{\text{inter-bs}}$).

For oral exposures, scaling doses by a fractional power of BW has been found to better account for interspecies differences in body size than scaling by BW as such. Since oral doses

are usually expressed as mg/kg BW, a correction factor is needed to convert the doses in mg/kg into allometrically scaled doses. Thus, the *DAF* and $AF_{\text{inter-bs}}$ are given by:

$$DAF_{\text{oral}} = (\text{animal BW}/\text{human BW})^{1-\alpha} \quad (3)$$

$$AF_{\text{inter-bs(oral)}} = (\text{human BW}/\text{animal BW})^{1-\alpha} \quad (4)$$

where α is the allometric power. This power is not exactly known, and can be represented by a distribution (e.g., normal). Because this adjustment is meant to extrapolate from the test animal to the median human, the average (median) animal BW in the study and the median human BW in the target (sub)population should ideally be used (U.S. EPA 2011b). If these are not available, then standard values can be used (e.g., U.S. EPA 1988), with an uncertainty that is probably negligible compared to the uncertainty in the allometric power (though the BW uncertainty could be included in the assessment).

For inhalation exposures, different types of *DAF*s have been derived for particles (Regional Deposited Dose Ratio, or RDDR) and gases (Regional Gas Dose Ratio, or RGDR) (U.S. EPA 1994). Based on interspecies information about respiratory tract geometries and air flow rates, the inhalation *DAF*s differ depending on whether the effects of interest are regional or systemic. For example, for effects in the upper airways, *DAF*s are based on the surface areas of relevant regions of the respiratory tract and the inhalation minute-volume. For systemic effects that involve transport by blood, *DAF*s utilize information on species differences (if any) in blood-air and blood-tissue partition coefficients. As with the oral *DAF*s, these are meant to extrapolate between the (median) test animal and the median human. Standard values, rather than statistically-based medians or values specific to the study, are usually employed, but clearly

these are uncertain as well. Thus, one could define the uncertainty in the *DAF* (or in the analogous $AF_{\text{inter-bs(inhalation)}}$) by assuming lognormal residual uncertainty:

$$DAF_{\text{inhalation}} = (\text{RDDR or RGDR}) \times \exp(\varepsilon_{DAF}) \quad (5)$$

$$AF_{\text{inter-bs(inhalation)}} = (\text{RDDR or RGDR})^{-1} \times \exp(\varepsilon_{DAF}) \quad (6)$$

where ε_{DAF} is normally distributed with a standard deviation of σ_{DAF}. The value of σ_{DAF} might be based on propagating the uncertainties in the parameters occurring in the calculations predicting the RDDR or RGDR or based on expert judgment.

Chemical-specific toxicokinetic or toxicodynamics differences

Test animals and human beings differ not only generically (e.g., in body size), but also in compound-specific toxicokinetics or toxicodynamics. Though on average across chemicals, the *DAF* is intended to neither under- nor overestimate the interspecies differences, the actual interspecies difference for any particular chemical is unknown in the absence of chemical-specific data. This uncertainty is addressed by subsequently dividing by a distribution for animal-to-human uncertainty (*AHU*) reflecting the additional differences in sensitivity between animal and median human beyond those addressed by the *DAF*, i.e., the toxicokinetic/dynamic differences specifically related to the chemical considered. For instance, assuming a lognormal uncertainty, one could define

$$AHU = \exp(\varepsilon_{AHU}) \quad (7)$$

where ε_{AHU} is normally distributed with a standard deviation of σ_{AHU}. With chemical-specific toxicokinetic or toxicodynamic data, a CSAF or DDEF may be developed, resulting in:

$$AHU = (\text{CSAF or DDEF}) \times \exp(\varepsilon_{AHU}) \quad (8)$$

where the standard deviation of ε_{AHU} would normally be smaller than that of the default value related to (6), as discussed in Supplemental Material, Chemical-specific/data-derived toxicokinetics or toxicodynamics.

Additional study-specific adjustments

Depending on the situation (e.g., experimental set-up of a critical study, toxicity database), additional issues may need to be addressed to infer the equipotent dose in the median human under the target conditions. Those additional adjustments and their associated uncertainties that are specific to the study (or endpoint) are addressed in Step 2 as well. The purpose is to account for these "other uncertainties" (*OU*) in characterizing the uncertainty distribution for the median human dose associated with a specified magnitude of effect, based on a specified study and endpoint. Examples of additional uncertainties include:

- The human hazard to be assessed relates to a different duration of exposure than that in the study. For instance, when the effect was in a subchronic rather than chronic study, the animal dose for the selected magnitude of effect might have been smaller in a chronic study. Based on historical data, one can estimate the empirical distribution for the ratio of chronic to subchronic dose (e.g., using equipotent doses from studies of both durations across many chemicals). Or, in specific situations a dose-time relationship (e.g., cumulative dose = constant, analogous to Haber's law) could be postulated, along with a distribution reflecting the uncertainty in how accurately the relationship holds.

- The human hazard to be assessed relates to a different route of exposure than that in the study, such as inhalation versus oral. Again, both an empirical (e.g., ratio of inhalation to oral equipotent doses), theoretical (e.g., based on total intake or absorbed dose), or model-based (e.g., physiologically-based pharmacokinetic [PBPK] model) approach can be used, along with a distribution reflecting the uncertainty in how accurately the assumed relationship is believed to hold.

- The hazard is being assessed for a different exposure pattern than that in the study, such as continuous exposure in humans vs. daily bolus exposure in the test animal. In this case, it is common to make assumptions about the dose-time relationship, such as peak or cumulative dose, as the basis for adjustment. If multiple assumptions are plausible, the uncertainty among the different options can be characterized through a distribution. For instance, when there is uncertainty if a given peak exposure would be equivalent to a three times lower or a three times higher equivalent continuous dose as compared to Haber's rule, this could be reflected by taking those values as the lower 5^{th} and upper 95^{th} percentiles of the equivalent dose distribution for constant exposure.

Note that in this step uncertainties are considered given the same magnitude of the same effect (endpoint). Uncertainties with respect to possibly different effects due to missing studies, even if they are at a similar level of severity, are not addressed here. This additional uncertainty is best addressed after completing Steps 1–3, which are all related to the specific effect under consideration. See Supplemental Material, Cross-study/endpoint uncertainties, for a discussion of some of these other additional uncertainties.

Step 3: Accounting for human interindividual variability in sensitivity

The aim of this step is to take into account differences in sensitivity across individuals in the population. For an exposure limit, for example, the result would be the uncertainty distribution for the dose associated with a specified endpoint and magnitude of effect (M*) for a "sensitive" individual, defined in terms of a percentile or incidence in the population (I*). To make these inferences, a population distribution representing the variation in equipotent doses among individuals needs to be specified. Because there are usually limited data as to the magnitude of this variation, this uncertainty needs to be taken into account as well.

Assuming a lognormal distribution for human variability, with standard deviation σ_H on a log-scale, the relationship between M*, the incidence $I_{\geq M^*}$ of effects greater than or equal to M*, and human dose HD is given by

$$I_{\geq M^*}(HD) = \Phi[\{\ln HD - \ln HD(0.5_{\geq M^*})\}/\sigma_H], \quad (9)$$

where Φ is the standard normal cumulative distribution. A similar relationship can be derived for any other assumed human variability distribution. For an exposure limit, one selects a target incidence value $I^*_{\geq M^*}$ and solves for dose D. Given that the median of the distribution $HD(0.5_{\geq M^*})$ was calculated in Step 2, this can be calculated by multiplying the median by the ratio between the I* quantile of the variability distribution and its median, denoted the human variability factor HV_{I^*}. For a lognormal distribution

$$HV_{I^*} = \exp\{z_{I^*}\,\sigma_H\}, \quad (10)$$

where z_{I^*} is the normal z-score corresponding to a quantile $I^*_{\geq M^*}$. For instance, at a 5% incidence, $z_{5\%} = -1.64$; at a 1% incidence, $z_{1\%} = -2.33$. Combining equations (2) and (10), the resulting equation is

$$HD(I^*_{\geq M^*}) = AD_{M^*} \times DAF \times HV_{I^*} / (AHU \times OU), \quad (11)$$

For discussion of recent analyses of human variability data, see IPCS (2014). For instance, Hattis and colleagues (Hattis et al. 2002; Hattis and Lynch 2007) estimated equipotent doses in a number of individuals, and calculated the standard deviations σ_H of the log-transformed equipotent doses, representing the variability in sensitivity among individuals. Then, they fitted a lognormal distribution to these standard deviations established for different chemicals (studies). They separated the available data into toxicokinetic and toxicodynamic factors, and estimated the uncertainty in the overall human variability as a combination of toxicokinetic and toxicodynamic variability. In this way, a default uncertainty distribution for intraspecies variation may be defined (IPCS 2014).

For some effects, we might suspect larger differences in sensitivity than others, or it might be known that the particular target subpopulation is highly sensitive for the agent considered. Or, we might be more uncertain for some effects than for others, for instance for effects that did not occur in the database underlying the default distribution. In such cases, one may decide to deviate from the default distribution in the appropriate direction. If compound- and endpoint-specific toxicokinetic or toxicodynamic data are available, these may be used to define a case-specific human variability distribution, with case-specific uncertainty about that distribution (discussed in Supplemental Material, Chemical-specific/data-derived toxicokinetics or toxicodynamics).

Monte Carlo calculation of HD_M^I

Keeping variability and uncertainty distinct in the calculation of HD_M^I requires a hierarchical approach to implementation. In addition, because the individual distributions cannot

be combined in closed-form (particularly incorporating uncertainty in the extent of human variability), a Monte Carlo simulation (MC) approach is necessary (See IPCS 2014, for an "approximate probabilistic approach" that can be implemented in a spreadsheet without Monte Carlo simulation). Specifically, at each MC iteration, all the steps addressing uncertainty are done first, followed by the steps evaluating variability:

- Evaluating uncertainty: Simultaneously draw MC samples [j] from AD_{M*}, DAF, AHU, OU, and σ_H. Obtaining MC samples from AD_{M*} is not a standard output from U.S. EPA's Benchmark Dose Software (BMDS) (U.S. EPA 2014b), but can be generated with PROAST using the bootstrap method (RIVM 2012). Bayesian methods offer another approach to generating such samples, and software such as WinBUGS, JAGS, or Stan can be used.

- Evaluating variability: Combine $(AD_{M*}[j] \times DAF[j])/(AHU[j] \times OU[j])$, to obtain one sample of the "median" human dose $HD(0.5_{\geq M*})[j]$. Next, given the target incidence $I*$, evaluate one sample of the human variability factor $HV_{I*}[j] = \exp\{z_{I*}\, \sigma_H[j]\}$. Combining these results is one MC sample of the human target dose associated with a particular incidence $I*$ and magnitude of effect $M*$: $HD_{M*}^{I*}[j] = HD(0.5_{\geq M*})[j] \times HV_{I*}[j]$.

- The result after many samples is the uncertainty distribution for HD_M^I.

For the illustrative datasets discussed below, this procedure was used with 10^7 MC samples for uncertainty (j) (for AD_{M*}, 10^3 bootstrap samples were resampled with replacement). Datasets and computer code are available in Supplemental Material (see Supplemental Material, Table S1

for details). An example of this approach for an exposure limit can be found in van der Voet and Slob (2007) who used the term I-CED rather than *HD*.

Calculating Population Incidence for Stochastic Quantal Endpoints

In the deterministic interpretation of quantal endpoints, the calculated incidence directly represents the expected incidence in the overall population. However, in the stochastic interpretation the calculated incidence relates to a single individual's probability *M* of experiencing the quantal endpoints (such as a tumor). For this reason, the HD_M^I for stochastic and deterministic quantal endpoints cannot be directly compared. To make such comparison possible, for the stochastic interpretation, the expected incidence in the overall population needs to be calculated by integrating all possible values of *M* (Slob et al. 2014). The calculation is simplified by the preceding assumption that the underlying continuous dose-response relationships are "monotonic and parallel on a log-dose scale across species and across individuals within a species." Specifically, let the animal dose-response function be represented by

$$M_A(AD) = f(AD, \theta) \tag{12}$$

where M_A is the magnitude of effect, *AD* is the animal dose, and *f* is the dose-response function with parameters θ. Based on "Step 2," the median human has the same magnitude of response as the animal [i.e., $M_A = M_{H,I>50\%}$] when the human dose $HD = AD \times DAF / (AHU \times OU)$. Rearranging so that $AD = HD \times AHU \times OU / DAF$, the dose-response function for the median human will be

$$M_{H,I>50\%}(HD) = f(HD \times AHU \times OU / DAF, \theta) \tag{13}$$

with the *same* model parameters θ. From "Step 3," the equipotent dose across human individuals is distributed lognormally with log-transformed standard deviation σ_H. Therefore, the magnitude of effect for a particular percentile of the population with z-score z, will be

$$M_{H,z}(HD) = f(\exp[z \times \sigma_H] \times HD \times AHU \times OU / DAF, \theta). \quad (14)$$

For a lognormally-distributed population of equipotent doses, z has a normal distribution. Therefore, the population arithmetic mean of M_H will be equal to the expected value of $M_{H,z}$ over a normally distributed z:

$$< M_H(HD) > = \int f(\exp[z \times \sigma_H] \times HD \times AHU \times OU / DAF, \theta)\ \varphi(z)\ dz \quad (15)$$

where $\varphi(z)$ is the standard normal probability density.

In the case of a stochastic quantal endpoint, M_H is the "individual probability of effect," which, averaged over the population in equation (15), would be, by definition, equal to the expected population incidence of effect. Uncertainties in the quantities θ, *DAF*, *AHU*, *OU*, and σ_H would then need to be propagated through the calculation to derive the uncertainty in this population incidence.

Monte Carlo calculation of population incidence for stochastic quantal endpoints

As with the prototypical implementation of the unified probabilistic framework described above, implementing the calculation of population incidence for a stochastic quantal endpoint, equation (15), requires a Monte Carlo (MC) simulation. As was the case for calculating HD_M^I, all the steps addressing uncertainty are done first, followed by the steps evaluating variability. In particular, at each value of human dose *HD* of interest:

- Evaluating uncertainty: Simultaneously draw MC samples [j] from θ, DAF, AHU, OU, and σ_H. Note that θ, which is generally multidimensional since most dose-response functions have more than one fitted parameter, has replaced the scalar (one-dimensional) quantity AD_{M^*} from above. Obtaining MC samples from θ is not a standard output from BMDS, but can be generated with PROAST using the bootstrap method. Bayesian methods offer another approach to generating such samples, and software such as WinBUGS, JAGS, or Stan can be used.

- Evaluating variability: Generate a human population by drawing N samples $z[k]$ from a standard normal distribution, and calculate the mean value over z of M_H:

$$< M_H(HD)>[j] = \Sigma_{k=1\ldots N} f(\exp[z[k] \times \sigma_H[j]] \times HD \times AHU[j] \times OU[j] / DAF[j], \theta[j]) / N \quad (16)$$

where N is large enough for convergence.

The result after many samples [j] is the uncertainty distribution for $< M_H(HD)>$. For a stochastic quantal endpoint, this then equals the expected population incidence of the quantal effect. This procedure was used for the stochastic quantal treatment of tumors with 10^7 MC samples for uncertainty (j) (for θ, 10^3 bootstrap samples were resampled with replacement) and 10^4 MC samples for variability (k). Datasets and computer code are available in Supplemental Material (see Supplemental Material, Table S1 for details).

Illustrative Datasets Analyzed

Two datasets contained as examples in the PROAST software (RIVM 2012) are used to illustrate the approach: body weight changes in rats and forestomach tumors in male mice. The tumor dataset is analyzed multiple ways – as deterministic quantal data, and as stochastic quantal data, and with extra risk levels of 10% and 1%. As discussed above, the HD_M^I outputs obtained for stochastic and deterministic quantal endpoints cannot be directly compared, so for the stochastic interpretation, the expected tumor incidence in the overall population is also calculated.

The uncertainty distributions for each step are based on the following:

- The uncertainty in AD_{M^*} (= BMD at BMR=M^*) is estimated via the bootstrap method in PROAST. To address model uncertainty, a standard set of models is fit, with the results of all models having goodness-of-fit p-values > 0.05 combined with equal weight.
- The distributions for *DAF* and *AHU* from Bokkers and Slob (2007), based on historical data on interspecies BMD ratios, are assumed:
    - *DAF* =(BW animal/BW human)$^{(1-\alpha)}$, with α assumed to have a normal distribution with mean 0.7 and standard deviation 0.024.
    - *AHU* has a lognormal distribution with geometric mean 1 and geometric standard deviation of 2.0. Notably, this distribution includes the current U.S. EPA default animal-to-human uncertainty factor (UF_A) of 3 applied after application of a deterministic *DAF* within its 95% confidence interval (U.S. EPA 1994; 2011a).

The combined distribution for the animal-to-human adjustment, when applied to rats or mice, includes the commonly-used animal-to-human factor of 10 within its 95% confidence interval.

- *OU* is omitted from the analysis (the critical study is assumed to be an adequate chronic study).

- The distribution for σ_H is based on a reanalysis by IPCS (2014) of published human TK (37 datasets) and TD data (26 datasets) compiled by Hattis and Lynch (2007). The result is a lognormal distribution for σ_H with geometric mean 0.746 and geometric standard deviation 1.59. The resulting distribution for the human variability factor HV_{I^*} (equation 10) when evaluated at an incidence $I^* \leq 5\%$, includes the commonly-used human variability factor of 10 within its 95% confidence interval.

Results

For each example dataset, the results of each step in the probabilistic approach are summarized in Table 4: (1) BMD modeling to estimate the animal dose-response relationship (see also Figure 4A and 5A), (2) probabilistic interspecies adjustments to estimate the equipotent doses in median humans (HD_M), and (3) probabilistic estimates of human variability to estimate the equipotent dose in sensitive humans (HD_M^I) (see also Figure 4B and Figures 5B-5D).

Representative BMD modeling results are shown in Figure 4A for the continuous dataset (body weight changes) and Figure 5A for the quantal dataset (tumors). For the body weight changes, the exponential and Hill models were fit, both of which had goodness-of-fit p-values > 0.05. For tumors, the multistage, Weibull, log-logistic, log-probit, gamma, and logistic models

were fit, of which only the multistage model failed to have a goodness-of-fit p-value > 0.05. These datasets show clear dose-responses and the uncertainty in the BMD is relatively modest, with confidence intervals (ratio of 95th percentile to the 5th percentile) ranging from 1.5- to 4.4-fold.

With respect to HD_M, the confidence intervals are wider due to the additional uncertainty in the interspecies adjustment (e.g., for Example A, 5.5/0.53=10 versus 11/5.6=2.0; see Table 4). Additionally, the 95th percentile of the HD_M is lower than that of the BMD, due to the allometric scaling factor.

With respect to HD_M^I, the confidence intervals are wider still, due to the additional uncertainty in intraspecies variability, and span a 40- to 60-fold range. The 5th percentile of the HD_M^I (underlined in Table 4, shown by the black square in Figure 4B and Figures 5B-5D) might be used as the "probabilistic RfD," interpreted as the lower (one-sided) 95% confidence limit on the dose at which an incidence of I^*=1% of the population experiences effects greater than the chosen critical effect size M^*. Note that the choice of percent confidence, critical effect size M^*, and the protection incidence I^* are informed by risk management considerations, and may depend on the specific context for the exposure limit. In the tumor example, a lower value for M^* (individual tumor risk) may be chosen by risk managers, even though it relates to only 1% of the population. Usually, however, risk managers may prefer to have an estimate of the expected tumor incidence in the overall population (which is worked out below).

Figures 4B and 5B-5D shows the 90% confidence intervals (i.e., 5th and 95th percentiles) for HD_M^I at difference levels of incidence I as a function of exposure, for a specified value of M^*. With this figure in hand, the different options as to protection incidence (in combination

with M^*) might be selected for deriving an exposure limit. The advantage of the probabilistic framework is illustrated by its transparency in the output: the magnitude of effect, the fraction of the population protected, and percent confidence, are all explicitly and quantitatively made visible. Moreover, uncertainties related to very small magnitudes of effect and/or very small incidences in the population can be made explicit and transparent (discussed in Supplemental Materials, Extrapolation to magnitudes of effect below a critical effect size and Extrapolation to very low incidences).

In the deterministic interpretation of the observed tumor incidence (example B, Table 4 and Figure 5B), the calculated incidence directly represents the expected incidence in the overall population. However, in the stochastic interpretation (examples C and D, Table 4 and Figures 5C-5D) the calculated incidence relates to a single individual's tumor probability M^*. Thus, the exposure limit in the deterministic case protects the relevant fraction of the population $(1 - I)$ against cancer as such, whereas the exposure limits in stochastic cases protect this fraction against the specified extra risk of cancer. For this reason, the outputs obtained from the deterministic versus the stochastic interpretation of tumor data cannot be directly compared.

As discussed in Methods, to compare the results from both interpretations, the expected tumor incidence in the overall population needs to be calculated for the stochastic interpretation, by integrating all the incidences I over all possible values of M. The results of this analysis, including uncertainty, are shown in Table 5 and Figure 6, where the confidence intervals on the population incidence of tumors are compared between the assumptions that tumors are "stochastic quantal" versus "deterministic quantal" effects. Results from a traditional linear extrapolation approach are also calculated for comparison.

These results clearly show that the confidence intervals for each of the two probabilistic approaches are wider than the difference between the confidence intervals (Figure 6 and Table 5). Therefore, at least in this example, the uncertainty in treating tumors as a deterministic versus a stochastic endpoint is not as great as the other uncertainties that have been characterized. Further, in this example, the result from a traditional linear extrapolation approach is not lower than the lower (one-sided) 95% confidence limit, so in that sense it is not necessarily "conservative" at the 95% level. The latter result was also found in various example cases examined by Slob et al. (2014).

Conclusions

Compared to previous probabilistic approaches to dose-response assessment (Baird et al. 1996; Evans et al. 2001; Gaylor and Kodel 2000; Hattis et al. 2002; Slob and Pieters 1998; Swartout et al. 1998), the framework proposed here is the first to unify across the various types of endpoints that may occur in toxicological studies, such as continuous vs. quantal endpoints, or cancer vs. noncancer endpoints. It does so by treating all endpoints as having a (direct or underlying) continuous response (at the level of an individual). It thereby fulfills the NRC (2009) suggestion to develop a unified approach to dose-response assessment for all endpoints. Furthermore, as discussed in Supplemental Material, the framework described here can incorporate other advances in toxicology and risk assessment, such as probabilistic exposure assessment (Supplemental Material, Integrating with probabilistic exposure assessment), CSAFs or DDEFs (Supplemental Material, Chemical-specific/data-derived toxicokinetics or toxicodynamics), and adverse outcome pathways (Supplemental Material, Extrapolation to downstream health endpoints and adverse outcome pathways and Figure S1).

The main idea of the framework proposed here is to quantify all relevant uncertainties by distributions instead of using conservative (single) values. However, for some uncertainties, it is currently unclear how to quantify them. Importantly, the uncertainty associated with the identification of the critical studies and endpoints is difficult to quantify, and the usual (deterministic) approach of focusing on the most sensitive studies and endpoints is hard to avoid. Consequently, even if the probabilistic approach described here is implemented, the result might be more conservative than it appears. For instance, if the particular species, strain, and sex of animal were idiosyncratic (the effect would not occur in humans) or particularly sensitive compared to humans, then the derived HD_M^I would be biased downwards. Furthermore, the most sensitive study from a large collection of studies will likely be more "conservative" than the most sensitive study from a smaller number of studies. The current approach remains unsatisfactory – be it in a deterministic or in a probabilistic assessment. In the short-term, the uncertainties related to the choice of the biological model might be better characterized by carrying forth multiple species/strains/sexes and endpoints to dose-response analysis (e.g., as recommended by NRC 2011), resulting in multiple HD_M^I estimates that reflect uncertainty in the chosen biological model. Furthermore, the emergence of studies using multi-strain rodent panels or genetically diverse population-based rodent models (as opposed to single, homogeneous, inbred strains) might provide a means to partially address these uncertainties quantitatively (e.g., Chiu et al. 2014; Rusyn et al. 2010).

Additionally, even conditional on the appropriate biological model, a number of implementation challenges remain. However, although these issues have become more apparent in developing the probabilistic framework, they are equally relevant for any (deterministic) dose-

response assessment method. The most important conceptual issue that has not yet been resolved is the question of which quantal endpoints should be treated as deterministic or stochastic quantal endpoints. While for histopathological quantal data, the deterministic interpretation is obvious from first principles, it is not directly clear whether cancer or malformation quantal data should be treated as stochastic or as deterministic quantal data (Slob et al. 2014). The problem is that it is not possible to directly establish this distinction from interpretation of single experiments, so additional research is needed as to what methods or datasets can distinguish between these options.

As a practical matter, there may be a tendency to treat more severe endpoints (such as tumor incidence) as stochastic since, at first sight, they seem to lead to more conservative results (although this may not always be the case). If, however, an endpoint is in reality deterministic rather than stochastic, then the outcome from the probabilistic dose-response assessment would be based on experimental error rather than biological phenomena. We repeat that this problem would not be specific for the probabilistic framework, but equally holds for traditional deterministic dose-response assessments, such as those that apply linear extrapolation.

Another conceptual issue related to stochastic quantal endpoints concerns the definition of toxicologically-equivalent effect metric for individual probability of effect (e.g., of malformations or cancer). Specifically, it remains unclear how individual probability of effect observed in animals can be made equivalent to individual probability of effect in humans in situations where background risks differ greatly between test animal and humans. Correction for background risk can be done in various ways, such as additional, extra, or relative risk, but there are no conclusive scientific arguments to favor one over the other.

Although the framework discussed here aims to estimate health effects in the human population in terms of both *M* and *I*, it is often practical to choose a specific value of *M* (or maybe several) to simplify the calculations, as well as the output. Therefore, the choice of the critical effect size *M** (or BMR) is often relevant. For continuous endpoints, current conventions as to the critical effect size *M** are based on a combination of biological considerations and statistical limitations of typical dose-response data (e.g., EFSA 2009). For instance, it was argued by EFSA (2009) that the transition from NOAEL to BMDL should not result in a systematic change in derived exposure limits in the long run, resulting in a recommended default for continuous endpoints of BMR = 5%. Of course, deviations in the default are allowed if biologically substantiated (e.g., BMR = 20% or more for liver enzyme levels, BMR = 10% for cholinesterase activity). Furthermore, one is reminded that the final output from the dose-response assessment includes the value of *M*, so that it remains visible. Consequently, one might consider to require a lower value for *I* if the value of *M* is suspected to be higher than desirable from a public health perspective. For deterministic quantal endpoints, the value of *M** is implicitly defined by the data (i.e., the associated severity category), though in some cases more than one category may be reported (e.g., "mild," "moderate," and "severe"). For stochastic (quantal) endpoints, *M** relates to the individual probability of effect (although, in this case, the overall population incidence can be calculated as well, in which case *M* vanishes).

Additionally, there is of course the issue of choosing values (i.e., uncertainty distributions) to be used as inputs in the probabilistic dose-response assessment. First, it should be noted that the uncertainty in the BMD is quantified by the BMD confidence interval. In the probabilistic framework, this uncertainty directly propagates through to the overall uncertainty in

the outcome of the dose-response assessment. In this way, it is directly visible to what extent designing more quantitatively informative experiments would improve a specific dose-response assessment, i.e., it might indicate that further improvement would substantially decrease the overall uncertainty in the HD_M^I , or that the impact would be minor. In terms of the adjustments from the POD, uncertainty distributions for particular aspects have been suggested based on meta-analyses of historical data (Bokkers and Slob 2005; Bokkers and Slob 2007; Hattis et al. 2002; Hattis and Lynch 2007), and reviewed by IPCS (2014). Thus, in those cases where no case-specific information for a given aspect is available, these distributions may be applied as a preliminary "default" distribution in probabilistic dose-response assessments. The historical data underlying these distributions were not generated for that purpose, and it might be argued that they are not always perfectly representative or highly informative. The fact that the probabilistic methodology exists makes it highly valuable to gather and/or generate data that may lead to better-supported uncertainty distributions. Therefore, further research and exploration of historical data that may inform the uncertainty distributions would be highly useful. One of the greatest challenges is a better characterization of human toxicodynamic variability, for which there are much less data as compared to toxicokinetic variability. Emerging molecular-biology and high-throughput systems, such as use of genetically diverse populations of human cells, offer some opportunities to address this data need in a more expedited fashion (Abdo et al., 2015; Zeise et al. 2013).

Furthermore, we note that issues in choosing input values holds equally for non-probabilistic dose-response assessments – the main difference is that the latter methods often use single default values, most of which have been generally accepted by the risk assessment

community, largely by convention. However, the current single default values lead to point estimates with unknown level of confidence and unspecified level of the protection of the population. By contrast, the probabilistic framework allows one to examine quantitatively the uncertainty, variability, and magnitude of effect associated with dose-response assessments using such conventional approaches. In the examples worked through here using the postulated uncertainty distributions, the result of default approaches, such as dividing an animal BMDL by 100 or linearly extrapolating from an allometrically-scaled animal BMDL, were higher than the (one-sided) 95% confidence limit of the probabilistic outputs for the protection goals, in terms of the magnitude of effect and population incidences illustrated. A similar result was found in case studies of carcinogens by Slob et al. (2014) when comparing with linear extrapolation. These results imply that these traditional deterministic approaches are not necessarily conservative in the sense that the derived "virtually safe" dose does not always reach 95% confidence.

Ultimately, as noted by the NRC (1996; 2009), a probabilistic framework will provide a substantially more complete quantitative characterization of hazard. In particular, in conjunction with exposure data, the relative impact of different risk management options – in terms of magnitude of effect, incidence in the population, and degree of confidence – will be much more explicit and transparent. It is envisioned that this will lead to better-informed risk management decisions.

References

Abdo N, Xia M, Brown CC, Kosyk O, Huang R, Sakamuru S, Zhou YH, Jack J, Gallins P, Xia K, Li Y, Chiu WA, Motsinger-Reif A, Austin CP, Tice RR, Rusyn I, Wright FA. 2015. Population-Based in Vitro Hazard and Concentration-Response Assessment of Chemicals: The 1000 Genomes High-Throughput Screening Study. Environ Health Perspect. doi: 10.1289/ehp.1408775.

Abt E, Rodricks JV, Levy JI, Zeise L, Burke TA. 2010. Science and decisions: advancing risk assessment. Risk Anal. 30(7):1028-1036.

Bailer AJ, Noble RB, Wheeler MW. 2005. Model uncertainty and risk estimation for experimental studies of quantal responses. Risk Anal. 25(2):291-299.

Baird SJS, Cohen JT, Graham JD, Shlyakter AI, Evans JS. 1996. Noncancer Risk Assessment: A Probabilistic Alternative to Current Practice, Human Ecolog. Risk Assess. 2(1):70-102.

Barlow S, Renwick AG, Kleiner J, Bridges JW, Busk L, Dybing E et al. 2006. Risk assessment of substances that are both genotoxic and carcinogenic report of an International Conference organized by EFSA and WHO with support of ILSI Europe. Food and Chemical Toxicology, 44: 1636-1650.

Benford D, Bolger PM, Carthew P, Coulet M, DiNovi M, Leblanc JC et al. 2010. Application of the Margin of Exposure (MOE) approach to substances in food that are genotoxic and carcinogenic. 2010. Food and Chemical Toxicology, 48:Suppl 1, S2-S24.

Bokkers BG, Slob W. 2005. A comparison of ratio distributions based on the NOAEL and the benchmark approach for subchronic-to-chronic extrapolation. Toxicol Sci. 85(2):1033-1040.

Bokkers BG, Slob W. 2007. Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Crit Rev Toxicol. 37(5):355-373.

Chiu WA, Campbell JL Jr, Clewell HJ 3rd, Zhou YH, Wright FA, Guyton KZ et al. 2014. Physiologically based pharmacokinetic (PBPK) modeling of interstrain variability in trichloroethylene metabolism in the mouse. Environ Health Perspect. 122(5):456-63. doi: 10.1289/ehp.1307623

Crump KS. 1984. A new method for determining allowable daily intakes. Fundam Appl Toxicol. 4(5):854-871.

Dedrick, RL. 1973. Animal scale-up. J Pharmacokinet Biopharm 1:435-461.

European Food Safety Authority (EFSA). 2005. Opinion of the Scientific Committee on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. The EFSA Journal, 282, 1-31.

European Food Safety Authority (EFSA). 2009. Guidance of the Scientific Committee on a request from EFSA on the use of the benchmark dose approach in risk assessment. The EFSA Journal, 1150, 1-72.

Evans JS, Rhomberg LR, Williams PL, Wilson AM, Baird SJ. 2001. Reproductive and developmental risks from ethylene oxide: a probabilistic characterization of possible regulatory thresholds. Risk Anal. 21(4):697-717.

Gaylor DW, Kodell RL. 2000. Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. Risk Anal. 20(2):245-250.

Hattis D, Baird S, Goble R. 2002. A straw man proposal for a quantitative definition of the RfD. Drug Chem Toxicol. 25(4):403-436.

Hattis D, Lynch MK. 2007. Empirically Observed Distributions of Pharmacokinetic and Pharmacodynamic Variability in Humans—Implications for the Derivation of Single Point Component Uncertainty Factors Providing Equivalent Protection as Existing RfDs. In *Toxicokinetics in Risk Assessment*, J. C. Lipscomb and E. V. Ohanian, eds., Informa Healthcare USA, Inc., 69-93.

International Program on Chemical Safety (IPCS). 2005. Chemical-Specific Adjustment Factors for Interspecies Differences and Human Variability: Guidance Document for Use of Data in Dose/Concentration-Response Assessment. World Health Organization, Geneva. Available: http://whqlibdoc.who.int/publications/2005/9241546786_eng.pdf [accessed 28 April 2015]

International Program on Chemical Safety (IPCS). 2014. Guidance Document on Evaluating and Expressing Uncertainty in Hazard Characterization. World Health Organization, Geneva. Available: http://www.who.int/ipcs/methods/harmonization/areas/hazard_assessment/en/ [accessed 28 April 2015]

Kleiber, M. 1932. Body size and metabolism. Hilgardia 6:315-353.

Kopylev L, Chen C, White P. 2007. Towards quantitative uncertainty assessment for cancer risks: central estimates and probability distributions of risk in dose-response modeling. Regul Toxicol Pharmacol. 49(3):203-207.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S et al. 2013. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol. 34(1):1-18.

National Research Council (NRC). 1996. Science and Judgement in Risk Assessment. National Academy Press: Washington, DC.

National Research Council (NRC). 2009. Science and Decisions: Advancing Risk Assessment. National Academy Press: Washington, DC.

National Research Council (NRC). 2011. Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde. National Academy Press: Washington, DC.

O'Brien J, Renwick AG, Constable A, Dybing E, Muller DJ, Schlatter J et al. 2006. Approaches to the risk assessment of genotoxic carcinogens in food: a critical appraisal. Food and Chemical Toxicol, 44: 1613-1635.

Pottenger LH, Becker RA, Moran EJ, Swenberg JA. 2011. Workshop report: identifying key issues underpinning the selection of linear or non-linear dose-response extrapolation for human health risk assessment of systemic toxicants. Regul Toxicol Pharmacol. 59(3):503-510.

Price PS, Keenan RE, Swartout JC. 2008. Characterizing interspecies uncertainty using data from studies of anti-neoplastic agents in animals and humans. Toxicol Appl Pharmacol. 233(1):64-70.

Rhomberg LR. 2011. Practical Risk Assessment and Management Issues Arising were we to Adopt Low-Dose Linearity for all Endpoints. Dose Response. 9(2):144-157.

Rhomberg LR, Goodman JE, Haber LT, Dourson M, Andersen ME, Klaunig JE et al. 2011. Linear low-dose extrapolation for noncancer health effects is the exception, not the rule. Crit Rev Toxicol. 41(1):1-19.

RIVM (National Institute of Public Health and the Environment). 2012. PROAST. Available: http://www.proast.nl [accessed 28 April 2015].

Rusyn I, Gatti DM, Wiltshire T, Kleeberger SR, Threadgill DW. 2010. Toxicogenetics: population-based testing of drug and chemical safety in mouse models. Pharmacogenomics. 11(8):1127-1136. doi: 10.2217/pgs.10.100.

Science Advisory Board (SAB). 2002. Review of Draft Trichloroethylene Health Risk Assessment: Synthesis and Characterization: An EPA Science Advisory Board Report. U.S. Environmental Protection Agency Science Advisory Board: Washington, DC. EPA-SAB-EHC-03-002. Available: http://yosemite.epa.gov/sab/sabproduct.nsf/D14C306CF5482E41852571CE00697543/$File/ehc03002.pdf [accessed 28 April 2015]

Shao K, Gift JS. 2013. Model Uncertainty and Bayesian Model Averaged Benchmark Dose Estimation for Continuous Data. Risk Anal. 34: 101–120. doi: 10.1111/risa.12078.

Slob W, Bakker MI, Te Biesebeek JD, Bokkers BGH. 2014. Exploring the uncertainties in cancer risk assessment using the integrated probabilistic risk assessment (IPRA) approach. Risk Anal 34(8):1401-1422.

Slob W, Pieters MN. 1998. A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. Risk Anal. 18(6):787-798.

Slob W, Setzer RW. 2014. Shape and steepness of toxicological dose-response relationships of continuous endpoints. Crit Rev Toxicol 44: 270-297.

Swartout JC, Price PS, Dourson ML, Carlson-Lynch HL, Keenan RE. 1998. A probabilistic framework for the reference dose (probabilistic RfD). Risk Anal. 18(3):271-282.

U.S. EPA. 1988. Recommendations for and documentation of biological values for use in risk assessment. (EPA/600/6-87/008). Cincinnati, OH: U.S. Environmental Protection Agency, National Center for Environmental Assessment. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=34855 [accessed 28 April 2015]

U.S. EPA. 1994. Methods for Derivation of Inhalation Reference Concentrations (RfCs) and Application of Inhalation Dosimetry. U.S. Environmental Protection Agency, Office of Research and Development, Office of Health and Environmental Assessment, Washington, DC, EPA/600/8-90/066F. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=71993 [accessed 28 April 2015]

U.S. EPA. 2005. Guidelines for Carcinogen Risk Assessment. U.S. Environmental Protection Agency, Washington, DC, EPA/630/P-03/001F. Available: http://www2.epa.gov/osa/guidelines-carcinogen-risk-assessment [accessed 28 April 2015]

U.S. EPA. 2011a. Recommended Use of Body Weight 3/4 as the Default Method in Derivation of the Oral Reference Dose. U.S. Environmental Protection Agency, Washington, DC, EPA/100/R11/0001. Available: http://www2.epa.gov/osa/recommended-use-body-weight-34-default-method-derivation-oral-reference-dose [accessed 28 April 2015]

U.S. EPA. 2011b. Exposure Factors Handbook 2011 Edition (Final). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-09/052F. Available: http://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=236252 [accessed 28 April 2015]

U.S. EPA. 2014a. Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors (DDEF) for Interspecies and Intraspecies Extrapolation. U.S. Environmental Protection Agency, Washington, DC, EPA/100/R-14/002F. Available: http://www2.epa.gov/osa/guidance-applying-quantitative-data-develop-data-derived-extrapolation-factors-interspecies-and [accessed 28 April 2015]

U.S. EPA. 2014b. Benchmark Dose Software (BMDS). Available: http://www.epa.gov/ncea/bmds/index.html [accessed 28 April 2015].

van der Voet H, Slob W. 2007. Integration of probabilistic exposure assessment and probabilistic hazard characterization. Risk Anal. 27(2):351-371.

Wheeler MW, Bailer AJ. 2007. Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Anal. 27(3):659-670.

Woodruff TJ, Sutton P. 2014. The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes. Environ Health Perspect. doi:10.1289/ehp.1307175.

Zeise L, Bois FY, Chiu WA, Hattis D, Rusyn I, Guyton KZ. 2013. Addressing human variability in next-generation human health risk assessments of environmental chemicals. Environ Health Perspect. 121(1):23-31.

Table 1. Components of adjustment, variability, and uncertainty in some typical Uncertainty Factors.

Uncertainty Factor	Adjustment	Variability	Uncertainty	Comment
Correcting dose for body size	✓		✓	Oral dose in mg/d may be adjusted to mg/(kg$^{\alpha}$ d), where the value of α may be chosen to be 1 or smaller than 1; this value is assumed to hold generically, so there is no variability, but the value of α is uncertain. Generic adjustments have also been derived for inhalation exposures based on regional gas or particle dosimetry derived from respiratory tract geometry and airflow.
Interspecies **TK or TD differences**	(✓)[a]		✓	Assuming that the test animal and humans are (on the appropriate dose scale) equally sensitive on average over chemicals, no further adjustment is needed (i.e., the factor equals one). But species do differ in sensitivity from one chemical to another. This chemical-to-chemical variability translates into uncertainty about the appropriate factor for a single chemical.
Intraspecies		✓	✓	It is expected that some humans will be more sensitive than others, but for a single chemical and effect, it is uncertain how many of them are more sensitive and by how much. So, there is variability, the size of which is uncertain.
Subchronic-chronic	✓		✓	On average over chemicals a given effect may be expected to occur at a lower dose with chronic exposure than with subchronic exposure (hence adjustment) but a single chemical may deviate to an uncertain degree.
Database	✓		✓	When one study type systematically results in lower PODs, then adjustment would be needed, while a single chemical may deviate to an uncertain degree.

[a] The adjustment factor is assumed to be one in this case, so that it appears to be absent in the calculations.

Table 2. Summary of Unified Probabilistic Framework.

Step and Goal	New Input(s) for Each Step	Output(s) for Each Step
1. Critical effect dose in animal. Estimate the uncertainty distribution for AD_{M*}, the animal dose associated with the critical effect size M*.	• Animal dose-response data • Toxicologically equivalent effect metric (M) • Critical effect size (M*) • Appropriate BMD analysis	AD_{M*} = uncertainty distribution for BMD based on analysis of animal dose-response data.
2. Equipotent dose in median human. Infer the uncertainty distribution for $HD_{M*} = HD(0.5_{\geq M*})$, the human dose at which 50% of the human population has effects greater than or equal to the critical effect size M*.	• AD_M* from Step 1. • DAF, distribution for the dosimetric adjustment factor due to differences in body size between animal and human. • AHU, distribution for the "animal-to-human uncertainties" due to unknown chemical+species-specific toxicokinetic or toxicodynamic differences. • OU , the distributions for "other uncertainties" due to study-/endpoint-specific conditions that differ from the target conditions.	$HD_{M*} = AD_{M*} \times DAF$ / ($AHU \times OU$) = uncertainty distribution derived by multiplying AD_{M*} by uncertain factors.
3. Equipotent dose in sensitive human (for an exposure limit). Infer $HD_{M*}^{I*} = HD(I*_{\geq M*})$, the dose at which a target incidence $I*_{\geq M*}$ yields effects of size $\geq M$*. Select a particular value HD* from the uncertainty distribution based on level of confidence.	• $HD(0.5_{\geq M*})$ from Step 2, serving as the uncertainty distribution for the median of the human variability distribution. • A lognormal human variability distribution, and a separate uncertainty distribution for its variance σ_H^2.[a] • A target incidence I*., from which a human variability factor HV_{I*} for the ratio between the "sensitive" and median individual is calculated [=exp($z_{I*}\ \sigma_H$) for a lognormal distribution, where z_I is the normal z-score for the I* quantile]	$HD_{M*}^{I*} = HD_{M*} \times HV_{I*}$ = uncertainty distribution for the I* percentile of a human variability distribution with median equal to HD_{M*} and standard deviation on log scale of σ_H.

[a]We use a lognormal distribution for the uncertainty in the variance, but other distributions could in principle be used.

Table 3. Example approaches to analysis of the animal dose-response data.

Endpoint type (examples)	***M*: Example toxicologically-equivalent effect metric.**	***M**: Example critical effect size(s)**	**Benchmark dose approach**
Continuous (hematocrit, serum enzyme, BW, organ/BW ratio)	Percent change relative to control	5%, 10% (percent change)	Continuous models with BMR= *M** = 5% , 10%.
Deterministic quantal (hepatic lesions, cytoxicity)	Severity category	"Minimal" (severity category)	Quantal models for 50% incidence of *M** = minimal, mild.
Stochastic quantal (hepatic tumors, fetal resorptions, eye malformations)	Extra risk for individual probability of occurrence	1%, 5%, 10% (extra risk)	Quantal models with BMR= *M** = 1%, 5%, 10%.

Table 4. Summary of Example Probabilistic Analyses. The two numbers between brackets represent 5th and 95th percentiles of the derived uncertainty distributions. All numbers representing dose are in mg/(kg d).

Example	**A**	**B**	**C**	**D**
Dataset	Body weight in rats[a]	Forestomach tumors in mice[b]	Forestomach tumors in mice[b]	Forestomach tumors in mice[b]
Type of endpoint	Continuous	Deterministic quantal	Stochastic quantal	Stochastic quantal
***M*: Effect Metric**	% change body weight	Tumor/No tumor	Individual probability (extra risk) of tumor	Individual probability (extra risk) of tumor
M: Critical effect size**	5%	Tumor[c]	10% extra risk	1% extra risk
***AD*$_{M^*}$: Critical effect dose (BMD) in chronic animal study**	(5.6, 11)	(5.1, 7.5)	(1.7, 3.7)	(0.39, 1.72)
***HD*$_{M^*}$: Equipotent dose in median human**	(0.53, 5.5)	(0.19, 1.9)	(0.076, 0.83)	(0.021, 0.33)
I: Target incidence protected**	1%	1%	1%	1%
***HD*$_{M^*}^{I^*}$: Equipotent dose in sensitive human[d]**	(<u>0.031</u>, 1.4)	(<u>0.011</u>, 0.47)	(<u>0.0044</u>, 0.21)	(<u>0.0013</u>, 0.079)

Note: Values rounded to 2 significant figures.

[a]Use control BW of 0.496 kg for *DAF*.

[b]Use standard BW of 0.03 kg for *DAF*.

[c]For the deterministic quantal treatment of tumors, benchmark dose analysis uses ED_{50}.

[d]"Sensitive human" is defined by the target incidence, here 1%. An exposure limit ("probabilistic RfD") can be based the 5th percentile of the derived uncertainty distribution of $HD_{M^*}^{I^*}$ (<u>underlined</u>), equivalent to a lower (one-sided) 95% confidence limit.

Table 5. Human dose at various specified tumor incidences estimated by linear extrapolation and by the probabilistic approach based on treating tumors as deterministic quantal versus stochastic quantal effects for the example tumor dataset.

Population tumor incidence for example tumor dataset	**Linear extrapolation from allometrically scaled BMDL [a] in mg/(kg d)**	**Human dose assuming deterministic quantal effect (5^{th} and 95^{th} percentiles) in mg/(kg d)**	**Human dose assuming stochastic quantal effect (5^{th} and 95^{th} percentiles) in mg/(kg d)**
5%	0.11	(0.029, 0.67)	(0.020, 0.37)
1%	0.022	(0.011, 0.47)	(0.0062, 0.17)
0.1%	0.0022	(0.0034, 0.33)	(0.0012, 0.078)
0.01%	0.00022	(0.0013, 0.25)	(0.00018, 0.040)

[a]Based on U.S. EPA (2005) default approach, where the point of departure is the lower (one-sided) 95% confidence limit on the benchmark dose at a 10% extra risk, scaled to a human equivalent by multiplying by $(BW_{animal}/BW_{human})^{0.25}$.

Figures Legends

Figure 1. Deterministic quantal endpoints: quantal responses reflecting incidences of a continuous response above/below a fixed cut point. The various dose-response lines in the upper panel reflect the (hypothetical) dose-responses of individual animals.

Figure 2. Stochastic quantal endpoints: quantal responses reflecting individual probability of effect. The dashed curves in upper and lower panel are the same, representing the (hypothetical) dose-response of the median animal. In the upper panel, the solid lines represent (hypothetical) individual dose-response curves. In the lower panel, the solid line reflects the expected value of the observed quantal response from the population of individual responses, which is less steep than the dose-response of the median animal.

Figure 3. Implementation of the unified probabilistic framework to derive the uncertainty distribution for HD_{M*}^{I*} and a corresponding probabilistic RfD. In step 1, benchmark dose analysis is used to derive the uncertainty distribution for AD_{M*}. In step 2, this distribution is combined with uncertainties in dosimetric adjustment, animal-to-human toxicokinetics and toxicodynamics, and other study-specific limitations, to derive the uncertainty distribution for HD_{M*}. In step 3, the distribution is further combined with the uncertainty in the human variability factor corresponding to the selected incidence I^* in the population to derive the uncertainty distribution for HD_{M*}^{I*}. The lower 95% (one-sided) confidence limit on HD_{M*}^{I*} can be chosen as the "probabilistic RfD" corresponding to the selected values of M^* and I^*. See Methods and Table 2 for additional details. This approach is illustrated with two example datasets, with results shown in Table 4 and Figures 4 and 5.

Figure 4. Results of analysis of example continuous dataset (rat body weight changes). Panel A: Representative benchmark dose modeling results using the Hill model with M^*=5% change; Panel B: Median estimate (solid line) and 5th and 95th percentile estimates (dashed and dotted lines, respectively) for the incidence (I) of effects of size *greater than* M^* (i.e., 5% change in body weight) as a function of population exposure (Dose), i.e, $I_{\geq M*}$(Dose). For reference, also shown are the "probabilistic RfD" corresponding to a 1% incidence of effects of size greater than

*M** at 95% (one-sided) confidence (solid square), the 90% (two-sided) confidence interval (CI) for the benchmark dose (gray shaded region), and a deterministic RfD equal to the BMDL/100 (vertical grey line).

Figure 5. Results of analysis of example quantal dataset (forestomach tumors in mice). Panel A: Representative benchmark dose modeling results using the Weibull model. Multiple benchmark dose estimates are shown, with the ED_{50} corresponding to *M**=tumor, and the BMD_{10} and BMD_{01} corresponding to *M**=10% and 1% extra risk, respectively. Panels B-D: Median estimate (solid line) and 5th and 95th percentile estimates (dashed and dotted lines) for the incidence (*I*) of effects of size *greater than M** as a function of population exposure (Dose), i.e, $I_{\geq M^*}$(Dose). In panel B, mouse forestomach tumors are treated as a deterministic quantal endpoint, where as in panels C and D, tumors are treated as a stochastic quantal endpoint (C: *M**=10% extra risk; D: *M**=1% extra risk). For reference, also shown in each panel are the "probabilistic RfD" corresponding to a 1% incidence of effects of size greater than *M** at 95% (one-sided) confidence (solid circle) and the 90% (two-sided) confidence interval (CI) for the benchmark dose (gray shaded region).

Figure 6. Comparison of estimated human population tumor incidences as a function of exposure (Dose) when treating tumors as a deterministic or a stochastic endpoint. Shown are the 90% (two-sided) confidence intervals for human population tumor incidence calculated from the probabilistic approach, depending on whether tumors in the example dataset are treated as deterministic (solid line) or stochastic (dashed line) quantal endpoints. For reference, also shown is the population tumor incidence derived using the default U.S. EPA method of linear extrapolation from a point of departure equal to the animal $BMDL_{10}$ allometrically-scaled by multiplying by $(BW_{animal}/BW_{human})^{0.25}$ (grey line).

Figure 1.

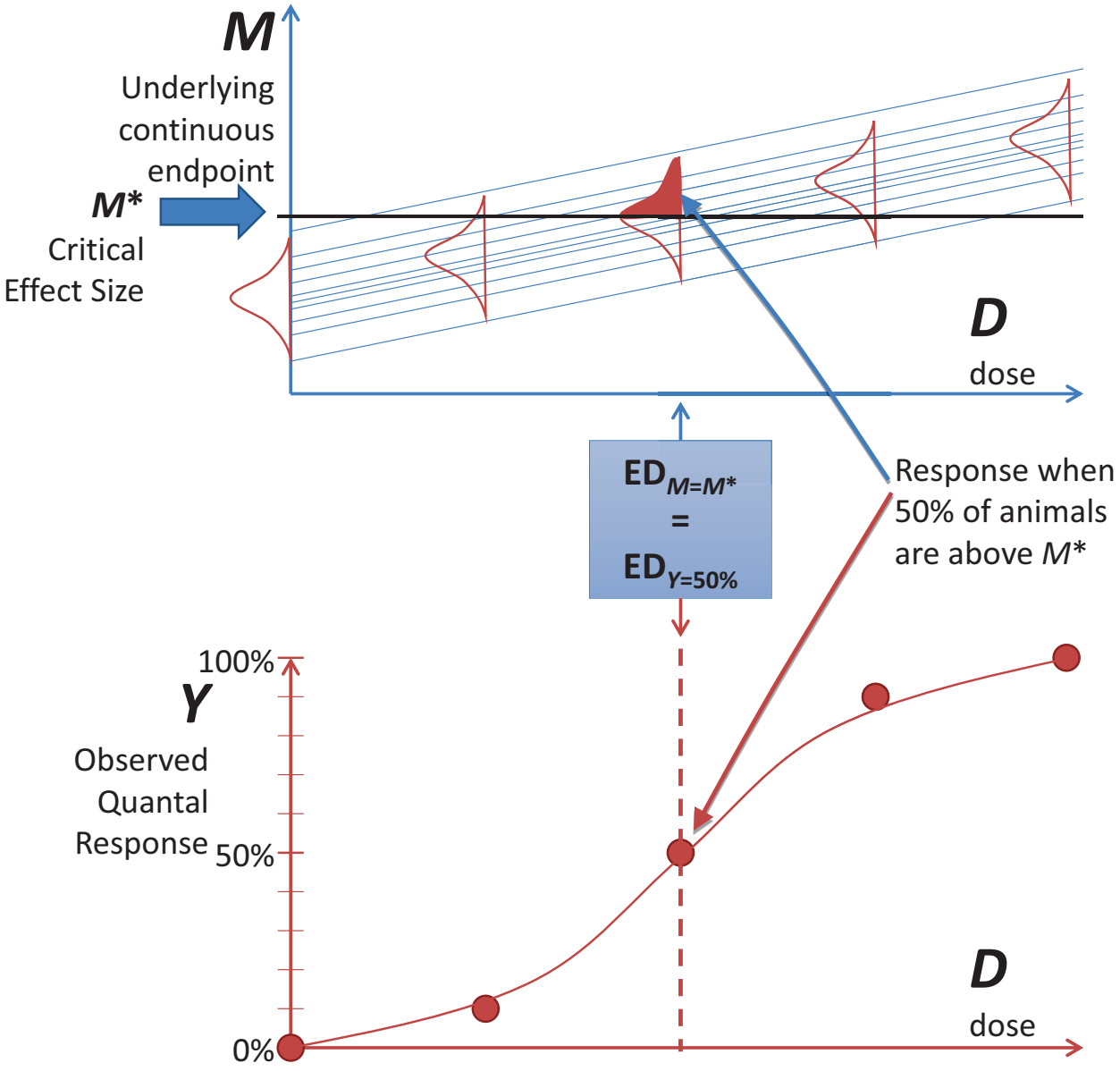
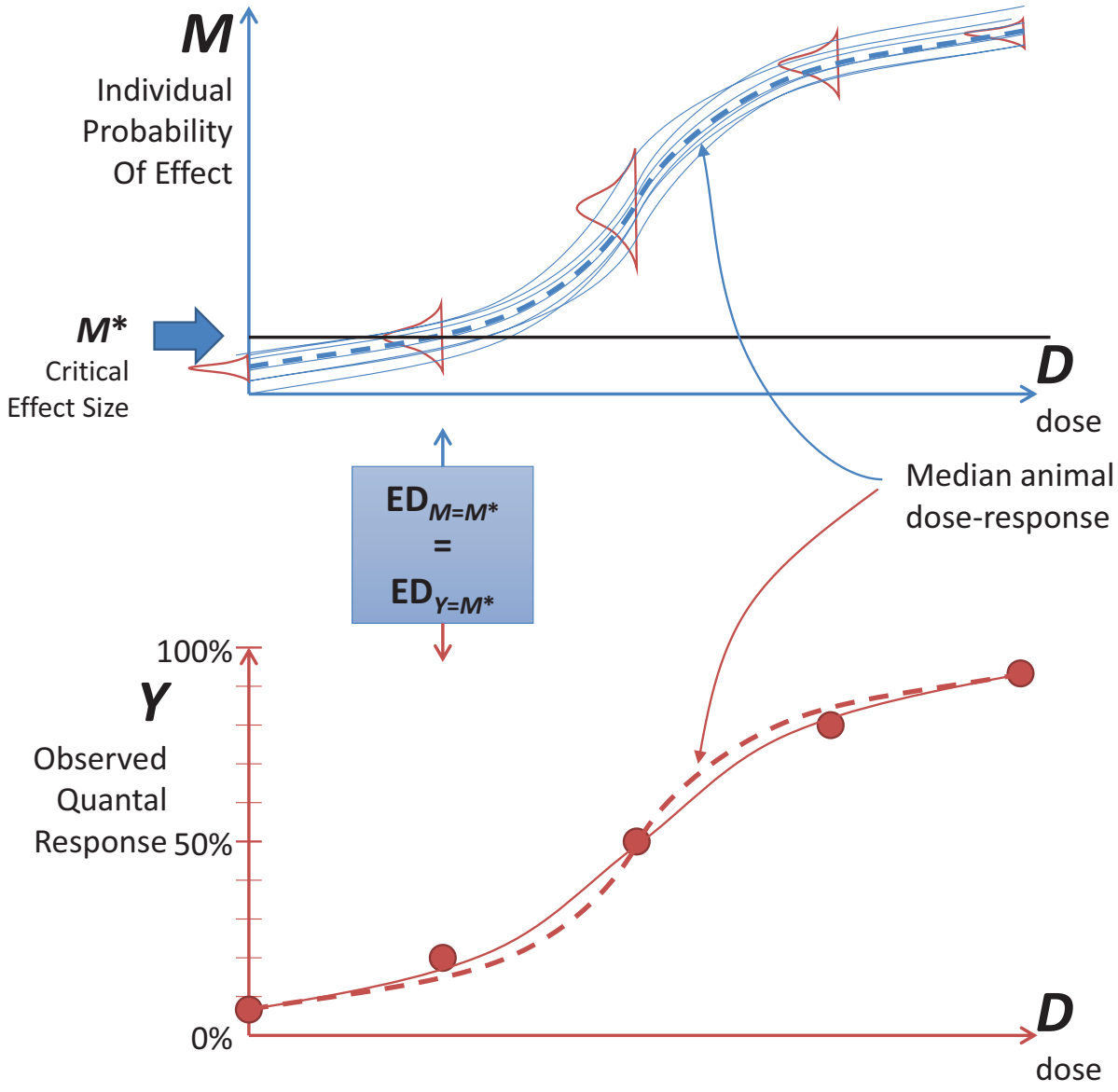

M
Underlying continuous endpoint
M*
Critical Effect Size
D
dose
ED M=M* = ED Y=50%
Response when 50% of animals are above M*
100%
Y
Observed Quantal Response
50%
0%
D
dose

Figure 2.


M
Individual
Probability
Of Effect
M*
Critical
Effect Size
D
dose
Median animal
dose-response
ED M=M*
=
ED Y=M*
100%
Y
Observed
Quantal
Response
50%
0%
D
dose

Environ Health Perspect DOI: 10.1289/ehp.1409385
Advance Publication: Not Copyedited

Figure 3.

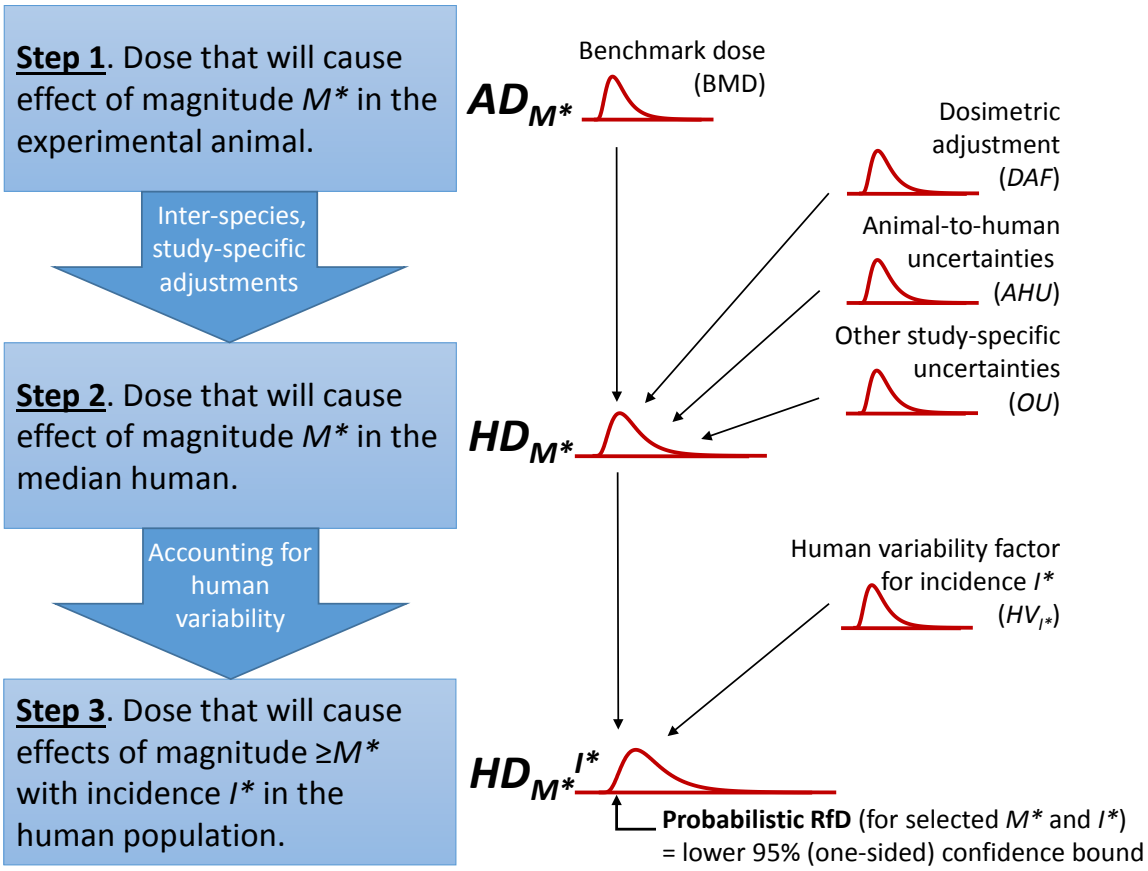


Figure 4.

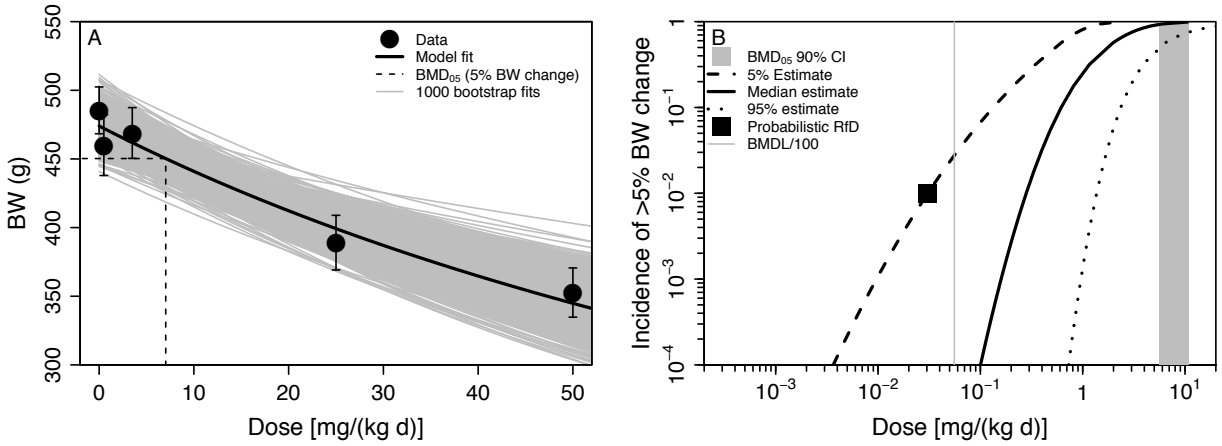

A
Data
Model fit
BMD05 (5% BW change)
1000 bootstrap fits
BW (g)
Dose [mg/(kg d)]
B
BMD05 90% CI
5% Estimate
Median estimate
95% estimate
Probabilistic RfD
BMDL/100
Incidence of >5% BW change
Dose [mg/(kg d)]

Figure 5.

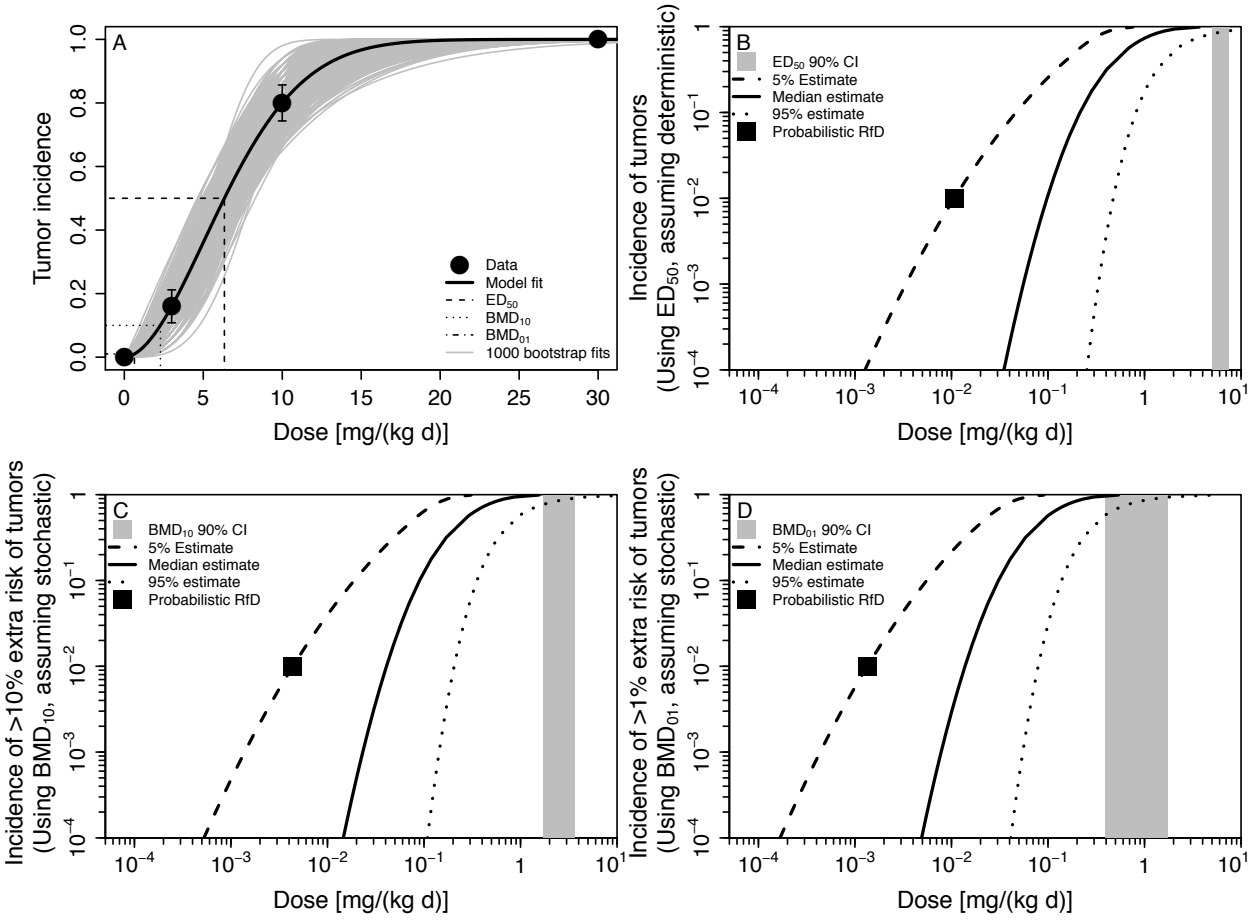
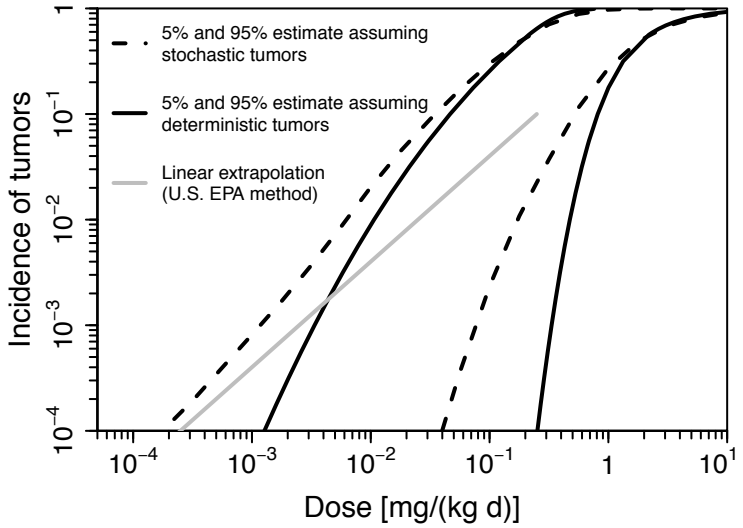

A
Tumor incidence
Dose [mg/(kg d)]
Data
Model fit
ED50
BMD10
BMD01
1000 bootstrap fits
B
Incidence of tumors
(Using ED50, assuming deterministic)
ED50 90% CI
5% Estimate
Median estimate
95% estimate
Probabilistic RfD
Dose [mg/(kg d)]
C
Incidence of >10% extra risk of tumors
(Using BMD10, assuming stochastic)
BMD10 90% CI
5% Estimate
Median estimate
95% estimate
Probabilistic RfD
Dose [mg/(kg d)]
D
Incidence of >1% extra risk of tumors
(Using BMD01, assuming stochastic)
BMD01 90% CI
5% Estimate
Median estimate
95% estimate
Probabilistic RfD
Dose [mg/(kg d)]

Figure 6.


5% and 95% estimate assuming stochastic tumors
5% and 95% estimate assuming deterministic tumors
Linear extrapolation (U.S. EPA method)
Incidence of tumors
Dose [mg/(kg d)]
10^{-4}
10^{-3}
10^{-2}
10^{-1}
1
10^{1}